JOURNAL OF APPLIED BEHAVIOR ANALYSIS 2008, **41,** 305–318 NUMBER 3 (FALL 2008)

BASELINE MEASUREMENT OF RUNNING AWAY AMONG YOUTH IN FOSTER CARE

LUANNE R. WITHERUP

KEYSTONE BEHAVIORAL PEDIATRICS

TIMOTHY R. VOLLMER, CAROLE M. VAN CAMP, AND HAN-LEONG GOH

UNIVERSITY OF FLORIDA

JOHN C. BORRERO

UNIVERSITY OF MARYLAND, BALTIMORE COUNTY

AND

KRISTIN MAYFIELD

ILEARN, INC.

The current study evaluated the use of various behavioral measures of running away with regard to (a) the differential utility of interval- versus event-based measures, (b) the differential utility of rate versus duration measures, (c) the utility of correcting for occurrence opportunity, and (d) the influence of unit of analysis (i.e., single-subject vs. grouped data). Seven different baseline measures were calculated for 84 runaways, and a unit-size analysis was conducted by constructing groups of various sizes from the original sample. An expert panel evaluated the suitability of the baseline measures for treatment evaluation. Results demonstrate the utility of evaluating duration-based measures and correcting for occurrence opportunity. Results also indicate that single-subject baselines may often be unacceptable for treatment evaluations, regardless of the type of measure selected for use.

DESCRIPTORS: baseline measures, foster care, running away

Running away is a severe form of problem behavior exhibited by adolescents (Biehal & Wade, 1999) that increases the likelihood of drug use and abuse (de Man, 2000; Edelbrock, 1980; Kennedy, 1991; Koopman, Rosario, & Rotheram-Borus, 1994; Yates, MacKenzie, Pennbridge, & Cohen, 1988), committing crimes (Abbey, Nicholas, & Bieber, 1997; Powers, Eckenrode, & Jaklitsch, 1990), engaging in prostitution (Cohen, MacKenzie, & Yates, 1991; Yates, MacKenzie, Pennbridge, & Swofford, 1991), contracting sexually transmitted diseases (Cohen et al.; Yates et al., 1991), attempting suicide (Kennedy; Powers et al.), joining street gangs (Yoder, Whitbeck, & Hoyt, 2003), skipping school (de Man; Sullivan & Knutson, 2000), and dropping out of school (Yates et al., 1991). Research also indicates that runaways are likely to be physically and sexually victimized while on the run (Abbey et al.; Hoyt, Ryan, & Cauce, 1999; Yates et al., 1991). It is important to note that these findings are correlational, preventing conclusions about the direction of causation or elimination of the possibility that a third variable could account for the relation between running away and increased risk of negative events. Nonetheless, there is enough cause for concern to warrant further research.

Given the serious risks listed above, several government reports and research studies have attempted to estimate the incidence of running away among youth in our society. Hammer, Finkelhor, and Sedlak (2002) estimated that in

Address correspondence to Luanne Witherup, 6867 Southpoint Dr. N, Suite 106, Jacksonville, Florida 32216 (e-mail: witherup@keystonebehavioral.com).

doi: 10.1901/jaba.2008.41-305

1999 approximately 1,682,900 children (2.6% of all U.S. youth) either ran away from home or were forced out by their caretakers (U.S. Bureau of the Census, 2000), a distinction that has been difficult to make in research studies. Foster children who run away have received considerable attention in recent years due to publicity surrounding children missing from substitute care and legal mandates for tracking missing children (see Florida Statute 937.022, 2004, as an example). Due to these legal mandates, estimates of running away among foster children are potentially more accurate than estimates for the general population, although considerable variability exists among estimates for foster children as well (Kaplan, 2004). Such variability is likely due to the use of varied definitions of running away, varied types of estimates (e.g., prevalence, incidence) and varied sampling procedures. For example, a report by the U.S. Department of Health and Human Services (2001) indicated that 9,112 foster children (2% of children in care) were on the run as of September 30, 2001. Fasulo, Cross, Mosley, and Leavey (2002) evaluated a specific sector of foster children, including 147 adolescents residing in specialized foster care, and found that 44% of the children ran away at least once and 22% ran away permanently. Estimates from studies that have examined the incidence of children exiting the child welfare system permanently by way of a run episode range from as low as 2% to as high as 21% (Courtney & Barth, 1996; U.S. Department of Health and Human Services).

Although incidence and prevalence estimates such as these are useful for understanding the breadth of this issue in our society, they do little to guide us in the assessment and treatment of individual children. Clinicians who work with runaways or potential runaways must obtain measures of the behavior at the level of the individual child to conduct a thorough assessment of the problem or properly evaluate the effects of an intervention. However, no investigations to date have attempted to obtain repeated measures of running away for individual children. Rather, researchers generally categorize children as either runaways or nonrunaways but make no attempt to the track the occurrence of run episodes on an individual basis.

The severity and relatively low rate of running away (e.g., a few times per year or month) present a unique challenge to behavior analysts with regard to repeated observation and measurement of environment–behavior relations as part of assessment and treatment evaluation. Repeated measures allow an analysis of functional relations and behavioral trends or patterns and are required for treatment evaluation using single-subject designs. For clinicians and researchers who study low-rate behavior such as running away, demonstrating treatment effectiveness in this way may prove difficult due to legal or ethical prohibitions against withholding or delaying intervention on the basis of an inadequate and highly variable baseline. One possible solution to this dilemma is to evaluate treatment effects across groups of individuals, an analytic strategy that has proven beneficial with other low-rate forms of behavior. For example, Agras, Jacob, and Lebedeck (1980) demonstrated the effectiveness of a community-wide water conservation intervention by using a multiple baseline design across cities.

The measurement of running away also presents a challenge with regard to determining the appropriateness of potential measures. Behavior is typically quantified across either time (interval based) or episodes (event based) and is recorded in terms of rate, duration, or interresponse time (IRT). There is currently no empirical basis for determining the most appropriate way to measure low-rate behavior, such as running away, even though the type of measure might affect the stability of behavioral trends.

Given the importance of measurement strategies for low-rate behavior and the chal-

lenges discussed above, the current study evaluated the use of various behavioral measures of running away. The analyses examined (a) the differential utility of interval- versus event-based measures, (b) the differential utility of rate versus duration measures, (c) the utility of correcting for occurrence opportunity, and (d) the influence of unit of analysis (i.e., single-subject, small-group, or large-group data). To evaluate the suitability of measures for treatment evaluation, a panel of expert judges was convened to evaluate the acceptability of individual and group baselines on the assumption that the baselines would eventually be used to evaluate the efficacy of a treatment.

METHOD

Inclusion Criteria and Demographics

Data for all runaway foster children residing in one Florida Department of Children and Families (FDCF) service district as of October 12, 2004, were considered for inclusion in this study. A runaway was defined as a child who engaged in one or more run episodes between September 1, 2001, and October 12, 2004. This time interval was deemed by FDCF personnel to represent the period of most accurate documentation of run episodes. Based on these criteria, 86 children were identified and 2 children were excluded due to missing or insufficient information. Of the remaining 84 runaways included in the analysis, 42 were female and 42 were male. The median age was 16 years (range, 10 to 17 years), the median number of run episodes was 2 (range, 1 to 19 episodes), the median number of days spent on the run was 10 (range, 1 to 441 days), and the median number of years spent in foster care was 2 (range, 0.12 to 15.6 years).

Data Collection

Behavior records took the form of missing-child reports, which are direct products of caregivers' responses to run episodes, rather than products of run episodes themselves, but are presumably correlated with actual run episodes. Data were obtained from two databases managed by FDCF. Data on run episodes initiated between September 1, 2001, and October 12, 2004, were obtained from the Missing Child Tracking System, which records the initiation and recovery dates of run episodes based on missing-child reports filed to the Florida Department of Law Enforcement. Demographic information including gender, age, and time spent in foster care was also obtained from the tracking system. A second database, HomeSafenet, was used to obtain the placement and removal dates for each placement episode experienced while in foster care and information about placements at lockdown facilities such as juvenile deten tion.

Data Analysis: Interval-Based Measures

Four interval-based baseline measures were calculated for each child across 30-day intervals beginning with the child's first day in care or September 1, 2001, whichever was later, and ending with the last completed interval expiring on or before October 12, 2004. The mean number of intervals evaluated for each child was 25 (range, 1 to 37 intervals).

Number of run initiations. The number of run initiations the child engaged in during each successive 30-day interval was calculated.

Proportion of opportunity days initiating a run. Run initiations cannot occur when a child is already on the run or placed in a lockdown facility, which may render number of run initiations inaccurate due to response restriction. Therefore, the number of opportunity days was calculated for each 30-day interval, with an *opportunity day* defined as any day not spent entirely on the run or in a lockdown facility. The number of opportunity days was then divided by the number of days in which a child initiated a run episode, resulting in a proportion of opportunity days with a run episode initiation to control for fluctuations in the number of opportunity days.

Number of days spent on the run. The total number of days spent on the run was calculated for each 30-day interval. Days in which a child spent at least some portion of the day on the run were considered to be a day spent on the run.

Proportion of opportunity days spent on the run. There is no opportunity to be on the run while in a lockdown facility, which may render number of days on the run inaccurate due to response restriction. Therefore, the number of opportunity days was calculated for each 30-day interval, with an *opportunity day* defined as any day not entirely spent in a lockdown facility (i.e., days with only a portion of the day spent in a lockdown facility were considered opportunity days). The proportion of opportunity days spent on the run was calculated for each successive 30-day interval by dividing the number of days spent on the run by the number of opportunity days.

Data Analysis: Episode-Based Measures

Based on an analysis of each child's run episodes, the following baseline measures were calculated using days as the unit of analysis.

Run durations. The duration of each run episode was calculated. Run episodes that were in progress on the date of data collection were indicated as such when displayed graphically. Therefore, minimum durations are depicted for such episodes because final durations remain unknown.

Episode IRT. The elapsed time between the end of each run episode and the beginning of the next episode was calculated. This measure was omitted for 29 children with only one run episode.

Initiation IRT. The elapsed time between successive run initiations was calculated. This measure was omitted for 29 children with only one run episode.

Data Analysis: Group-Size Analysis

Eighty of the 84 runners were randomly selected for inclusion in a group-size analysis to determine the usefulness of each measure for multiple baseline analyses across groups. A parametric group-size analysis was accomplished by constructing 31 groups as follows: 16 groups of 5, 8 groups of 10, 4 groups of 20, 2 groups of 40, and 1 group of 80. Each runner was randomly assigned to groups to approximate what often occurs in applied settings (i.e., intervention is implemented at a particular facility or region and not others). Although research on interventions targeted at the most highly recidivistic runners in particular is appealing, these treatment effects can likely be demonstrated on an individual rather than group basis due to the high rate of the behavior; thus, a randomization approach was used for this analysis.

Only the interval-based measures were subjected to the group-size analysis because the episode measures did not have a constant *x*-axis time progression. Baseline measures were calculated for each successive 30-day interval included in the span of the study (37 intervals total), and each data point represents the mean value for all runners within the group. The baseline lengths varied across runners, but the final interval graphed includes all runners (see inclusion criteria). Therefore, earlier group intervals represent progressively fewer individual runners due to varying durations in care. All groups contained at least 1 individual with 37 intervals of data. This aggregation method was chosen based on typical procedures employed by intervention researchers when examining aggregate data, but it does increase the likelihood that greater variability will be observed in earlier intervals compared to later intervals due to smaller sample size.

Expert Panel Evaluation

Similar to previous studies on data interpretation, a panel of 5 expert judges was constructed to evaluate the acceptability of each baseline measure for potential use during hypothetical treatment evaluations (Hagopian, Fisher, Thompson, & Owen-DeSchryver,

1997; Kahng et al., 1998). The judges reviewed both the single-subject and group data sets for all measures except initiation IRT. Initiation IRT was not reviewed because the direction of behavior change associated with improvement for the measure is ambiguous. For example, an improvement in the rate of running would produce an increase in this measure, but an improvement in the duration of run episodes would produce no change in this measure. So although analysis of initiation IRT may aid assessment by providing useful information about temporal patterning of run initiations, such a measure would not be appropriate for experimental evaluations of behavior change.

Selection. Five individuals were asked to serve on the expert panel based on their expertise in the field of applied behavior analysis and their experience working with runaway children, and all agreed to participate. Each expert judge possessed a doctorate degree, was a board-certified behavior analyst, had at least one first-author publication in the *Journal of Applied Behavior Analysis*, and had work experience with children who run away. Only 2 of the 5 individuals studied under the same faculty adviser, with 3 individuals receiving their degrees at the University of Florida, 1 from West Virginia University, and 1 from Louisiana State University.

Materials. Expert judges were asked to complete the evaluation independently at their leisure and were provided a maximum of 2 weeks to complete the task. In addition to a basic description of each measure as described above, judges were provided with written instructions (full text available from the first author) that specified that certain socially important problem behaviors are low rate (e.g., rape, murder, suicide, running away from home), and that it may be difficult to obtain adequate baseline measures for such behaviors. The experts were informed that the authors had compiled several relevant baseline measures in both single-subject and group formats using a sample of 84 foster children who had run away at least once. The experts were instructed to evaluate these data based on their expertise in the field of applied behavior analysis and experience working with children who run away, with the assumption that their role was that of a behavior analyst planning to evaluate an intervention for running away. The experts were asked to select a portion of the baselines for a proper experimental evaluation (e.g., a multiple baseline evaluation) based on the adequacy of the baselines for evaluating the intervention.

A total of 599 graphs were presented with interval-based single-subject measures presented first (336 graphs), followed by episode-based single-subject measures (139 graphs), and interval-based group measures (124 graphs). Each page included baselines of a single type (e.g., number of run initiations) for several unnumbered individual runners or groups of runners with group size denoted next to each group. Baseline measures were ordered randomly within each grouping of graphs to limit potential fatigue effects, and an analysis of disagreement rates across these successive sections did not reveal an upward trend in disagreements across graphs.

Data analysis. To evaluate the likelihood of baseline acceptance, the total number of baselines designated as acceptable by a majority of the expert panel (i.e., at least 3 of the 5 experts) was calculated for each runner individually. Note that runners could attain a maximum of six acceptable baselines (i.e., all measures except for initiation IRTs).

To evaluate possible differences in the likelihood of baseline acceptance based on type of measure and group size, the mean proportion of expert acceptance was calculated for each interval-based and episode-based measure individually (excluding initiation IRTs). Episode IRTs that were omitted for runners with only one run episode were automatically designated as inadequate (i.e.,

proportion expert acceptance = 0). The proportion of experts designating the baseline as acceptable was first calculated for each of the baseline graphs individually. The mean of these values was then calculated according to group size (interval-based measures only) and type of measure (e.g., number of run initiations, days spent on the run).

Interobserver Agreement and Interrater Agreement

Interobserver agreement for the calculation of all baseline measures was evaluated for 27 of the 84 children (32%) by having a second observer calculate the interval- and episode-based measures. For each measure, an agreement was scored if the two observers scored exactly the same measure in a given interval, and a disagreement was scored if the two observer scorings differed in any way for an interval. The mean agreement was then calculated for each type of measure by dividing the total number of intervals with agreements by the total number of agreements plus disagreements, then converting this ratio to a percentage. The mean agreement across all baseline measures was 99% (range, 98% to 100%).

Interrater agreement for the acceptability determinations of the expert panel was determined using a pairwise exact agreement comparison for the 599 acceptability ratings of each judge. Individual pairwise agreement scores were obtained by comparing each expert's ratings (i.e., acceptable or not acceptable) with the ratings of each other observer (5 observers = 10 pairings) across all graphs. The mean pairwise agreement score across all 10 pairings was 81% (range, 68% to 90%). A mean pairwise agreement score was also obtained for each observer. For example, if Observer 1 agreed with Observer 2 on 100% of the graphs but agreed with the remaining observers (i.e., 3, 4, and 5) on only 70% of graphs, the resulting mean pairwise agreement score for Observer 1 would be 77.5%. The mean pairwise agreement scores for the 5 observers were 72%, 82%, 82%, 84%, and 85%.

RESULTS

Single-Subject Interval-Based Measures

Although the utility of the four interval-based measures varied across children, three useful findings emerged as illustrated by the interval-based measures of 3 of the runners depicted in Figure 1. Each row shows the four interval-based measures for a given runner across successive 30-day intervals. The number of data points for each individual varies based on the amount of time spent in foster care; however, the final interval on each graph represents the same time period because all children were in care on the date of data extraction. The distribution of intervals along the x axis was adjusted for each child rather than retaining consistency across children because all comparisons were within subject rather than across subjects. Missing data points occur when an interval contained no opportunity days. Shading designates baselines judged to be acceptable for treatment evaluation by at least 3 of the 5 experts.

The first general finding that emerged from this analysis was that a majority of the children engaged in very few run episodes and in episodes of minimal duration (51% of children ran less than three times and spent less than 16 days on the run). Runner R28 (top row of Figure 1) illustrates this pattern. Similar information about behavioral trends was provided by all interval measures, and these baselines were typically judged to be unacceptable by the expert panel.

The second finding that emerged was that correcting for the opportunity to initiate a run and measuring duration of run episodes rather than rate proved useful. The rate of run initiations was not an appropriate measure for children with long run durations because being on the run artificially suppressed the opportunity to initiate new runs and deflated the rate measures. For such runners, improvements in the rate of run initiations were often accompanied by an increase in the amount of time spent

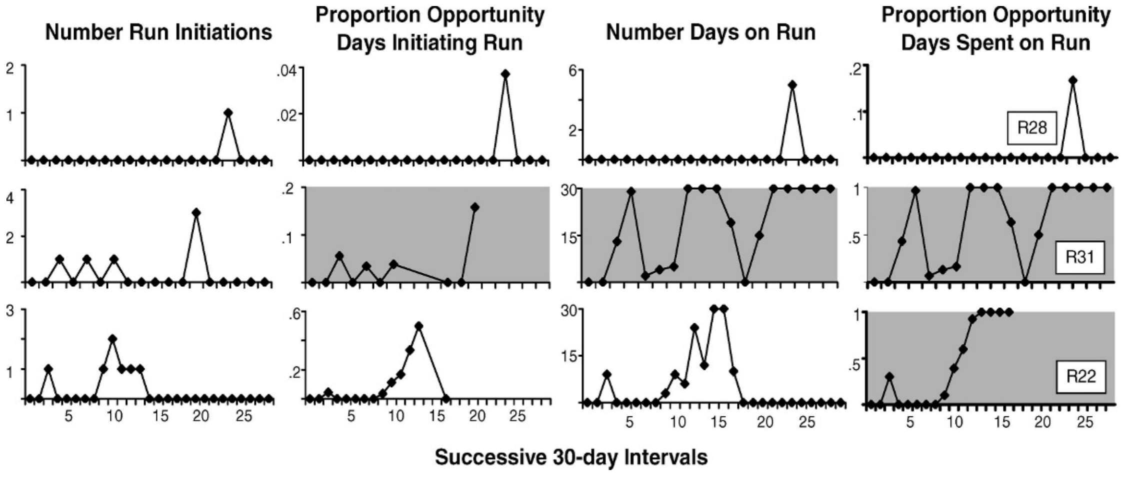

Figure 1. Example of single-subject interval-based measures. Each row represents data for 1 runner. The four interval-based measures are depicted in each column, with the *y*-axis labels included across the top as column headings. All measures are depicted across successive 30-day intervals on the *x* axis.

on the run (e.g., Runner R31, middle row of Figure 1, spent 327 days on the run). This runner showed a decline in number of run initiations (Column 1), but this decline did not represent a desirable outcome once the initiation opportunities were corrected (Column 2). For runners such as R31, the duration measures (Columns 3 and 4) provided the most accurate account given the substantial amount of time spent on the run. Correcting for initiation opportunity or measuring duration rather then rate proved to be particularly useful, in that total time spent on the run increased for a given runner and produced baselines that were more likely to be judged as acceptable for treatment evaluation by the expert panel.

The third general finding that emerged was the need to correct for days with no opportunity to run for children who spent a substantial amount of time in lockdown facilities. The data for Runner R22 (bottom row of Figure 1), who spent 402 days in lockdown, illustrate this effect. Although this runner showed a recent decline in the number of days spent on the run (Column 3), correcting for opportunity (Column 4) indicates that this was a forced improvement due to the significant amount of time spent in lockdown. By contrast, correcting for opportunity did not yield different information for Runner R31, who spent no time in lockdown. Correcting for opportunity to run for runners with significant lockdown histories produced baselines that were more likely to be judged as acceptable for treatment evaluation by the expert panel.

Single-Subject Episode-Based Measures

The episode-based measures allowed an explicit analysis of response duration and IRTs that were not possible using interval-based measures; however, the total number of run episodes affected the usefulness of these measures. An analysis of trend in run duration was possible only for children who engaged in two or more run episodes, and an analysis of IRT trend was possible only for children who engaged in three or more run episodes. Each row of Figure 2 depicts all three episode-based measures for 1 of 5 runners with common behavioral patterns. Successive run episodes are illustrated along the *x* axis; therefore, the number of data points for each child varied based on the total number of run episodes and IRT graphs were not applicable for runners with one episode. Scales of the *y* axis were adjusted on an individual basis to allow proper analysis

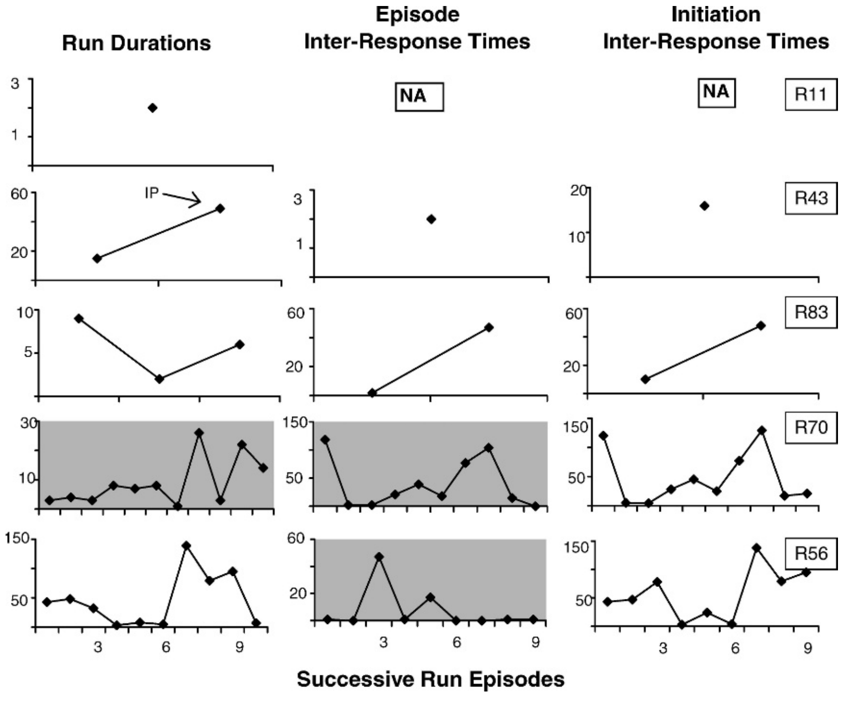

Figure 2. Example of single-subject episode-based measures. Each row depicts data for 1 runner. The three episode-based measures are depicted in each column, with the *y*-axis labels included at the top as column headings. All measures are depicted across successive run episodes on the *x* axis.

of trends. Active run episodes as of the date of data collection are designated as in progress (IP) and represent the minimum episode duration. Shading designates baselines judged to be acceptable for treatment evaluation by a majority of the expert panel. However, recall that initiation IRT baselines (last column) were not included in the expert panel evaluation.

Data for Runners R11, R43, and R83 in the top three rows are typical for children with few run episodes. Although data such as these provided limited information, it is important to note that with respect to run duration, limited information may still prove useful. For example, the fact that R11 remained on the run for only 2 days suggests the possibility that she may have been incapable of obtaining the basic needs required to maintain long absences from care (i.e., food, shelter). Such information could have important implications for treatment.

Episode-based measures for children who engaged in many run episodes were inherently more informative. For example, data for Runners R70 and R56 are much more descriptive due to the high number of run episodes. In general, differences between episode IRT (Column 2) and initiation IRT (Column 3) emerged for runners with relatively long run episodes. For example, the two measures are similar for Runner R83, who had a maximum run duration of 9 days, but differ substantially for Runner R56, who had a maximum run duration of 139 days. In general, an analysis of IRT measures may allow the identification of important functional relations if observed patterns are found to correlate with changes in other environmental conditions.

Expert Panel Single-Subject Evaluation

One question of interest is the likelihood that a given runner would have one or more

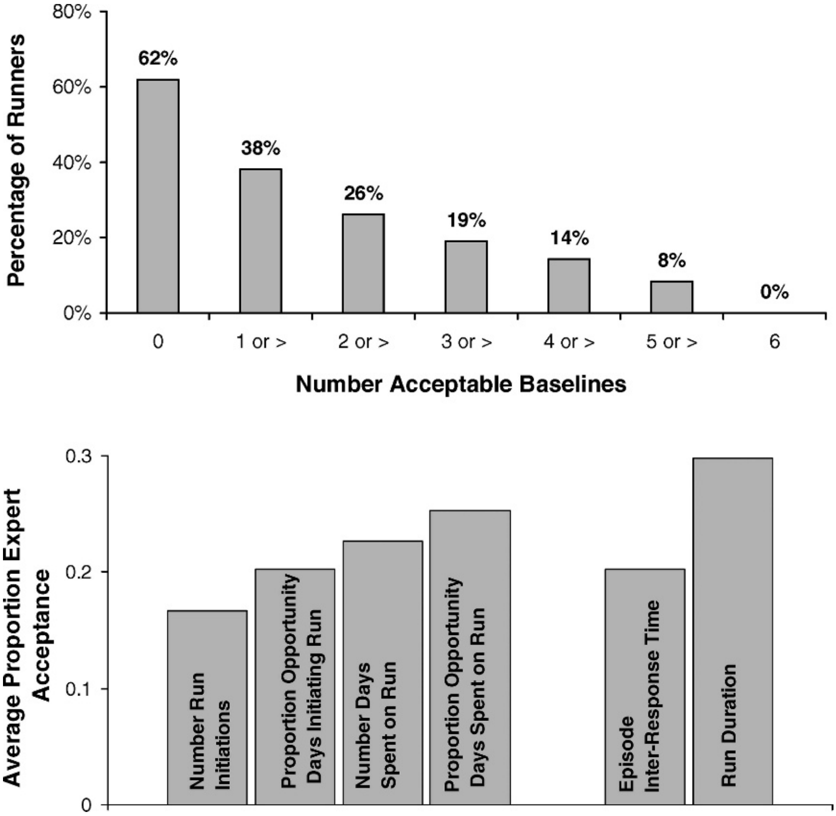

Figure 3. Top: expert panel evaluation (baseline acceptance). Baselines considered acceptable by a majority of the expert panel. Bottom: expert panel evaluation (measure type). The mean proportion of expert acceptance for all single-subject baseline measures.

baselines judged to be acceptable (samples illustrated by the gray shading in Figures 1 and 2). The upper panel of Figure 3 depicts the percentage of runners (*y* axis) with varying numbers of baselines judged to be acceptable by a majority of the expert panel (*x* axis). Because the expert panel did not evaluate initiation IRT graphs, a maximum of six acceptable baselines was attainable. Results indicate that a large percentage of runners (62%) had no baselines judged to be acceptable by the majority. The remaining 38% of the runners had at least one acceptable baseline, and none of the runners had all six baseline measures judged to be acceptable.

A second question of interest was whether the likelihood of baseline acceptance would vary according to the type of baseline measure selected. The lower panel of Figure 3 depicts the mean proportion of acceptance by the experts for all six baseline measures. The actual proportion of experts who accepted each baseline graph was determined, and then the mean of these values was calculated for each type of measure. Episode IRT baselines for children with only one run episode were automatically considered unacceptable because it is not possible to calculate the measure. Results for the interval-based measures (left side of graph) indicate that number of run initiations was the least accepted type of baseline measure (0.17), followed by the proportion of opportunity days initiating a run (0.20), the number of days spent on the run (0.23), and the

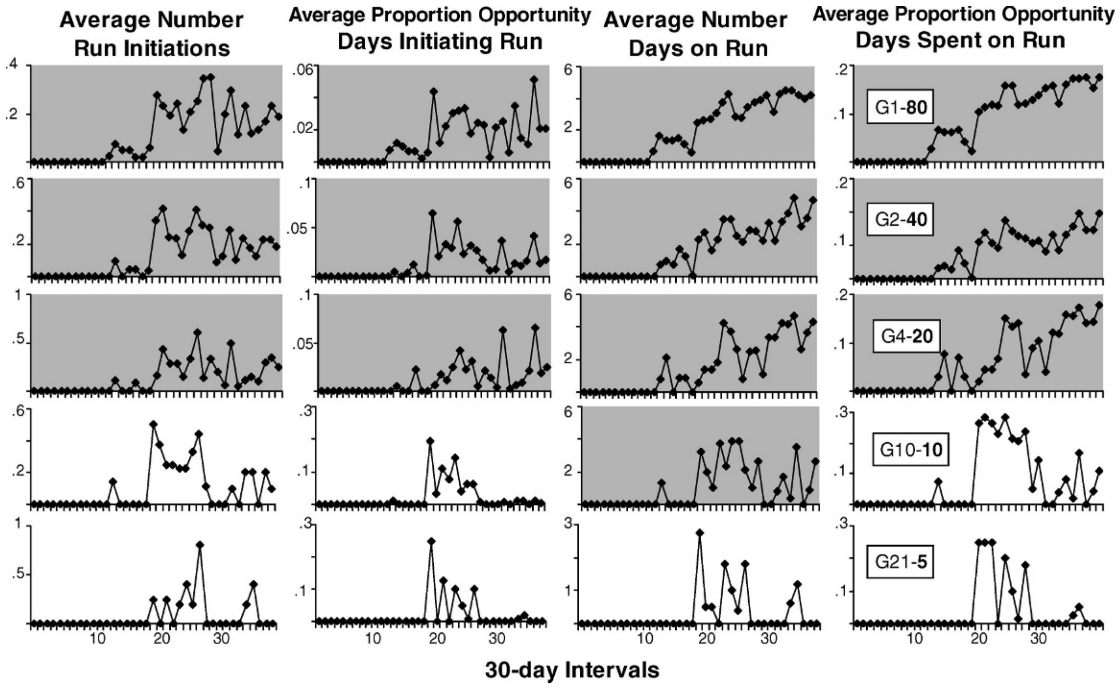

Figure 4. Example of group interval-based measures. Each row represents data for one group. Group sizes are in bold next to each group number. The four interval-based measures are depicted in each column, with the *y*-axis labels included across the top as column headings. All measures (group mean) are depicted across successive 30-day intervals on the *x* axis.

proportion of opportunity days spent on the run (0.25). Therefore, initiation measures (i.e., number of run initiations and proportion of opportunity days initiating a run) were less accepted than duration measures (i.e., number of days spent on the run and proportion of opportunity days spent on the run), and correcting for opportunity increased mean acceptance for both types of measures. Episode-based measures are depicted on the right side of the graph. Episode IRTs attained a mean acceptance similar to that of the interval-based measures (0.20), and run durations attained the highest acceptance overall (0.30). However, it is important to note that although successive run durations attained the highest overall acceptance, the suitability of this measure for treatment evaluation is inherently limited due the fact that it depends on the occurrence of the target behavior.

Group-Size Analysis

Each row of Figure 4 depicts all four interval-based measures for one of the five different-sized groups, with shading designating baselines judged to be acceptable by a majority of the expert panel. The group mean of each measure is depicted for each successive 30-day interval, with each data point representing the mean across all individuals who had data for that interval. Thus, the number of baselines included in the mean for a given interval varies, and the final interval depicts the mean of the final interval for all individuals contained within the group.

Although the shaded graphs in Figure 4 provide examples of acceptable group baselines, a more detailed parametric analysis of the degree to which group size would increase baseline acceptance was also conducted. Figure 5 depicts the mean proportion of expert

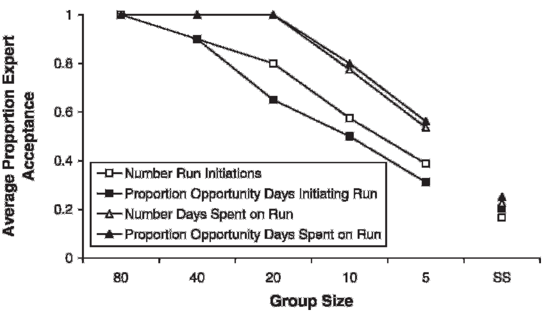

Figure 5. Expert panel evaluation (group-size analysis). The mean proportion expert acceptance is depicted according to group size (*x* axis) and type of measure (legend).

acceptance according to both type of measure (legend) and group size (*x* axis). Single-subject results were included for comparison (interval-based measures only).

Not surprisingly, results indicated that mean acceptance increased as the size of the group increased. This is likely due to the fact that aggregating data for multiple individuals allows more behavior to be captured and reduces variability in the data. Consistent with the single-subject analysis, duration measures fared better than or equal to initiation measures across all group sizes. Although maximum acceptance was reached by Group Size 20 for duration measures, initiation measures did not reach maximum acceptance until Group Size 80. One unexpected finding is also worth noting. Correcting for initiation opportunity (i.e., proportion of opportunity days initiating a run) did not produce any increases in acceptance, as was observed in the single-subject analysis. In fact, correcting for initiation opportunity actually decreased acceptance for Group Sizes 5, 10, and 20.

DISCUSSION

Given the lack of behavioral research targeted at the problem of running away, even the most basic issue of measurement has yet to be thoroughly addressed. Difficulties surrounding how and what to measure with respect to running away must be resolved before more complex issues such as the identification of behavioral function and treatment development can be addressed. The current investigation examined several different behavioral measures of running away and evaluated their utility for assessment and treatment evaluation. In general, the duration of run episodes rather than rate of occurrence is the more appropriate measure, and correcting for occurrence opportunity is beneficial particularly for children with lengthy run durations and extensive lockdown histories. Episode-based measures, including run durations and IRTs, were useful only for children with multiple run episodes and have limited usefulness for treatment evaluation. Except for highly recidivistic runners, single-subject baselines for all measures may often be unacceptable for treatment evaluations; thus, clinicians or researchers attempting to demonstrate treatment effects may need to evaluate groups of children using multiple baseline designs across groups with groups of at least 20 or more.

Although many individual baselines proved to be unacceptable for evaluation of treatment effects, each type of measure may have the potential to provide useful information when used in conjunction with other information. For example, additional assessment by a behavior analyst might reveal that Runner R28 (see Figure 1, top row) was separated from her siblings during the interval that contained her only run episode, which might lead to a function-based prevention strategy based on the hypothesis that separation from siblings serves as the primary establishing operation. Although all types of measures may be informative in some respect, results of this study highlight the need for both clinicians and researchers to carefully consider the possible implications of the type of measure they choose to use. Arbitrary selection of a measure could obscure pertinent information and ultimately hinder

treatment effectiveness or undermine the detection of important treatment outcomes.

Although research studies often focus only on the number of run episodes in a specified period of time (Hammer et al., 2002), these findings highlight the importance of avoiding rate-based measures for running away. Rate-based measures can be misleading due to the potentially long duration of the behavior during which new instances of behavior are not possible. This problem can be corrected by eliminating periods of ongoing behavior from the denominator of the rate calculation (i.e., the corrected initiation measure) or by using a duration-based measure rather than a rate measure. One could argue that duration measures are most appropriate for behavior such as running away in which a reduction in the duration of the behavior is a desirable outcome even if the rate of the behavior remains unchanged. For example, effects of recovery efforts for children already on the run may primarily be detectable in duration-based measures (i.e., run durations decrease with no effect on run initiations).

Another consideration illustrated by the results of this study is the need to correct for occurrence opportunity when measuring running away (i.e., proportion of opportunity days spent on the run). Whether using rate- or duration-based interval measures, it is important to account for periods of time in which environmental circumstances prevent the occurrence of a behavior. Failure to correct for occurrence opportunity may distort baseline data and alter interpretations regarding behavior change, particularly for children with substantial lockdown histories.

The relatively small percentage of expert-judged acceptable baselines indicates that experimental treatment evaluation may present significant difficulties for clinicians and researchers with all but the most recidivistic runners. Due to ethical concerns with reversal designs or intentional baseline extensions, the use of naturally occurring baselines appears to be the most promising approach to treatment evaluation. However, most individual baselines were not sufficient to demonstrate a convincing behavior change. Episode-based measures generally had the highest acceptability ratings, but a duration-based interval measure such as time spent on the run, which was the next most highly accepted measure when corrected for opportunity, might be most useful given that episode measures are dependent on the occurrence of the behavior to be eliminated.

Results of the expert panel evaluation suggest that grouping runaways in the context of single-subject methodology logic (e.g., multiple baseline across groups) may prove to be an effective strategy for treatment evaluation. Although baseline acceptability in the present study increased over that of single subjects for all group sizes, including as few as 5 runaways per group, results indicate that the use of duration-based interval measures with groups of 20 or greater would be the most effective approach. This strategy would allow behavioral researchers to conduct treatment evaluations for running away without abandoning single-subject design logic or being forced to rely on anecdotal report of treatment effectiveness. Nonetheless, the importance of simultaneously examining treatment effects on individual subjects should not be overlooked. More specifically, the ability to demonstrate the effects of a given intervention with even a single runaway can help add strength to demonstrations made at the group level.

At least three primary limitations to this study are worthy of note. First, the reliability of the data contained in the FDCF databases was not explicitly examined, although reporting and data-entry errors are almost inevitable. Future studies that mine data from large databases should include procedures for identifying and correcting such errors. Second, the data-aggregation method used in the group-size analysis resulted in the fewest number of runners in the early intervals and the largest number of runners in the later intervals, due to the varied amount

of baseline data available for each runner in the group. Although this method is likely to be used by clinical organizations and researchers who evaluate a group intervention, it inherently produces greater variability in earlier intervals (i.e., 1 to 12) due to the smaller sample size in those intervals. In actual treatment evaluations, researchers should use an inclusion criterion or limit the number of intervals evaluated so that all intervals represent a majority of the runners in the group. Finally, this investigation focused exclusively on measurement of running away using mined (i.e., collected retroactively) data rather than evaluation of any specific ongoing assessment or treatment procedures. However, these findings might serve as a catalyst for future research on behavioral assessment and treatment methods with this important population. For example, our own research team is currently investigating topics such as (a) child characteristics associated with running away, (b) maintaining variables for running, and (c) run probability by placement (e.g., group homes) and individual caregiver characteristics. Prevention strategies also warrant investigation, although behavioral researchers will face another methodological challenge in doing so because traditional single-subject research methods are not readily suited for an analysis of preventive interventions.

REFERENCES

Abbey, A. A., Nicholas, K. B., & Bieber, S. L. (1997). Predicting runaways upon admission to an adolescent treatment center. *Residential Treatment for Children & Youth*, *15*(2), 73–86.

Agras, W. S., Jacob, R. G., & Lebedeck, M. (1980). The California drought: A quasi-experimental analysis of social policy. *Journal of Applied Behavior Analysis*, *13*, 561–570.

Biehal, N., & Wade, J. (1999). Taking a chance? The risks associated with going missing from substitute care. *Child Abuse Review*, *8*(6), 366–376.

Cohen, E., MacKenzie, R. G., & Yates, G. L. (1991). HEADSS, a psychosocial risk assessment instrument: Implications for designing effective intervention programs for runaway youth. *Journal of Adolescent Health*, *12*(7), 539–544.

Courtney, M. E., & Barth, R. P. (1996). Pathways of older adolescents out of foster care: Implications for independent living services. *Social Work*, *41*, 75–83.

de Man, A. F. (2000). Predictors of adolescent running away behavior. *Social Behavior & Personality*, *28*(3), 261–268.

Edelbrock, C. (1980). Running away from home: Incidence and correlates among children and youth referred for mental health services. *Journal of Family Issues*, *1*(2), 210–228.

Fasulo, S. J., Cross, T. P., Mosley, P., & Leavey, J. (2002). Adolescent runaway behavior in specialized foster care. *Children & Youth Services Review*, *24*(8), 623–640.

Florida Statute 937.022. (2004).

Hagopian, L. P., Fisher, W. W., Thompson, R. H., & Owen-DeSchryver, J. (1997). Toward the development of structured criteria for interpretation of functional analysis data. *Journal of Applied Behavior Analysis*, *30*, 313–326.

Hammer, H., Finkelhor, D., & Sedlak, A. J. (2002). *National incidence studies of missing, abducted, runaway, and thrownaway children: Runaways/thrownaway children: National estimates and characteristics*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Hoyt, D. R., Ryan, K. D., & Cauce, A. M. (1999). Personal victimization in a high-risk environment: Homeless and runaway adolescents. *Journal of Research in Crime & Delinquency*, *36*, 371–392.

Kahng, S., Iwata, B. A., Fischer, S. M., Page, T. J., Treadwell, K. R. H., Williams, D. E., et al. (1998). Temporal distributions of problem behavior based on scatter plot analysis. *Journal of Applied Behavior Analysis*, *31*, 593–604.

Kaplan, C. (2004). *Children missing from care: An issue brief.* Washington, DC: Child Welfare League of America.

Kennedy, M. R. (1991). Homeless and runaway youth mental health issues: No access to the system. *Journal of Adolescent Health*, *12*, 576–579.

Koopman, C., Rosario, M., & Rotheram-Borus, M. J. (1994). Alcohol and drug use and sexual behaviors placing runaways at risk for HIV infection. *Addictive Behaviors*, *19*, 95–103.

Powers, J. L., Eckenrode, J., & Jaklitsch, B. (1990). Maltreatment among runaway and homeless youth. *Child Abuse & Neglect*, *14*, 87–98.

Sullivan, P. M., & Knutson, J. F. (2000). The prevalence of disabilities and maltreatment among runaway children. *Child Abuse & Neglect*, *24*, 1275–1288.

U.S. Bureau of the Census. (2000). *United States census 2000*. Retrieved January 12, 2006, from http://quickfacts.census.gov/qfd/states/12000.html

U.S. Department of Health and Human Services. (2001). *Adoption and foster care analysis and reporting system (AFCARS)*. Retrieved March 2, 2006, from http://www.aff.dhhs.gov/programs/cb/stats/tarreport/rpt0100/ar0100.html

Yates, G. L., MacKenzie, R., Pennbridge, J., & Cohen, E. (1988). A risk profile comparison of runaway and nonrunaway youth. *American Journal of Public Health*, *78*, 820–821.

Yates, G. L., MacKenzie, R. G., Pennbridge, J., & Swofford, A. (1991). A risk profile comparison of homeless youth involved in prostitution and homeless youth not involved. *Journal of Adolescent Health*, *12*(7), 545–548.

Yoder, K. A., Whitbeck, L. B., & Hoyt, D. R. (2003). Gang involvement and membership among homeless and runaway youth. *Youth & Society*, *34*(4), 441–467.

Received March 29, 2007
Final acceptance September 12, 2007
Action Editor, Linda LeBlanc